# The primary structure of protein L17 from the *Escherichia coli* ribosome

Wilfried Rombauts, Valère Feytons and Brigitte Wittmann-Liebold*

*Afdeling Biochemie, Departement Humane Biologie, Katholieke Universiteit, 49, Herestraat, 3000 Leuven, Belgium* and **Max-Planck-Institut für Molekulare Genetik, Abteilung Wittmann, Ihnestrasse 63–73, 1000 Berlin 33 (Dahlem), Germany*

Received 20 September 1982

The complete sequence of protein L17 which is a component of the large subunit of the *E. coli* ribosome has been determined. Peptides deriving from enzymatic hydrolysis with trypsin, thermolysin, chymotrypsin and *S. aureus* and *A. mellea* protease were isolated and sequenced by the DABITC/PITC double coupling method. Some overlapping peptides were obtained after mild acid cleavage of the protein. According to the amino acid sequence protein L17 contains 127 residues and has a molecular mass of 14 365. The primary structure of protein L17 agrees well with the amino acid analysis of the intact protein and its N-terminal sequence as derived from automatic sequencing in an improved Beckman sequencer. Secondary predictions and a search for homologous sequence stretches to other ribosomal proteins were made.

E. coli *ribosome* *Protein L17* *Primary structure determination*
*Secondary structure predictions* *Homology to other ribosomal proteins*

## 1. INTRODUCTION

Protein L17 is a component of the *E. coli* large ribosomal subunit [1] with $M_r$ = 14 000–17 000 as determined by SDS gel electrophoresis and hydrodynamic methods [2]. According to immune electron microscopy [3,4] it is located on the lower rear side of the hemispherically shaped large subunit. From cross-linking studies [5] and the analysis of protein complexes [6] it can be concluded that it is close to 8 other proteins within the 50 S particle. Protein L17 binds specifically to the 23 S RNA [7] and is incorporated early during the 50 S assembly process [8]. It is one of the most reactive proteins toward modification of the large subunit with *N*-ethylmaleimide [9,10], and as an isolated protein it is rather resistant to proteolytic degradation [11].

## 2. MATERIALS AND METHODS

Protein L17 was isolated [12] from 50 S ribosomal subunits of *E. coli* K12, strain A19, and the purity was checked by two-dimensional gel electrophoresis [13]. The pure protein was kindly provided by Dr H.G. Wittmann.

Peptide sets were generated by cleaving the protein with trypsin, chymotrypsin and thermolysin, under conditions described in [14]. Digestion with *Staphylococcus aureus* protease was performed in 50 mM ammonium bicarbonate buffer, pH 7.8 at 37°C for 18–24 h at an enzyme: substrate ratio of 1 : 30, and similarly in 50 mM ammonium acetate buffer (pH 4.0). Partial cleavage in dilute hydrochloric acid (10 mM, pH 2.0) was carried out at 108°C in intervals from 4–24 h, and the progress of the reaction was followed by removing aliquots and determination of the liberated aspartic acid residues. Some of the larger fragments of this cleavage were further split with trypsin or chymotrypsin.

In all these reactions, 100 nmol (1.5–2 mg) of L17 were used and the resulting peptides separated on cellulose thin-layer plates by two-dimensional electrophoresis/ascending chromatography, essentially as in [14]. About 10 nmol peptide mixture were applied per plate. In a later stage of the sequence determination HPLC-techniques were used


*Published by Elsevier Biomedical Press*
001457593/82/0000—0000/$2.75 

for the separation of some of the hydrolysates using reversed phase chromatography on RP-8 and RP-18 columns with dilute ammonium formate/methanol gradients [15].

Cyanogen bromide cleavage was carried out on 10 mg protein, dissolved in 2 ml 70% formic acid containing 2% 2-mercaptoethanol. The protein was reacted with 10 mg CNBr for 48 h, at room temperature in the dark. The resulting fragments were partially separated on Sephadex G-25 (superfine) in 0.1 N acetic acid and detected with ninhydrin, after alkaline hydrolysis.

A protease from *Armillaria mellea* [16] was used to cleave the protein at the N-terminal side of lysine residues. The enzyme was a kind gift from Dr V. Barkholt-Pedersen (Copenhagen). The digestion was carried out for 6 h at 37°C at pH 8 in 0.1 M *N*-methylmorpholine acetate, at an enzyme: substrate ratio of 1 : 1000. The detection and isolation of peptides containing the single cysteine residue of L17 was greatly facilitated by using *S*-[$^{14}$C]carboxamidomethylated L17 in some digestions. The *S*-carboxamidomethylcysteine-containing spots on the thin-layer plates were directly visualized by autoradiography. The reaction was done at pH 8.5 on 6.4 mg protein using iodo-[$^{14}$C]acetamide 53 $\mu$Ci/$\mu$mol (Amersham); the reaction conditions were as in [17]. Digestions with carboxypeptidase A and B were performed as in [14].

N-terminal residues of the peptides were determined by the dansyl technique, and their sequences by the DABITC/PITC (4-*N*,*N*-dimethylaminoazobenzene-4′-isothiocyanate/phenylisothiocyanate) double coupling method [18]. An improved Beckman sequencer [19,20] was used to degrade the intact protein.

Peptides (2–3 nmol) were hydrolysed in 200 $\mu$l 5.7 N HCl (Suprapur Merck) (containing 0.02% mercaptoethanol and 0.25% phenol) at 110°C for 20 h, under $N_2$, and the hydrolysate analysed on a Biotronik LC 2000 analyser, set at a sensitivity of 0.2 A. Total protein hydrolysates (100 $\mu$g sample) were prepared in a similar way, but hydrolysed for 20, 48, 72 and 96 h. The analyses were done at the same sensitivity. Cysteine was determined after performic acid oxidation and the absence of tryptophan verified by hydrolysis in 3 M mercaptoethane sulfonic acid as in [21].

## 3. RESULTS AND DISCUSSION

### 3.1. *Sequence determination*

The complete sequence of protein L17 (fig. 1) was derived from a combination of the partial or complete sequences of the peptides obtained from the cleavages with trypsin, *Staphylococcus aureus* protease, *Armillaria mellea* protease and dilute acetic acid. Results obtained from thermolytic and chymotryptic cleavages and from cyanogen bromide treatment confirmed certain sequences and overlaps. The use of *Armillaria mellea* protease was very helpful in establishing the final order of tryptic peptides, in particular for the region from position 40 to the C-terminus. Fig. 1 shows the order of the various peptides and overlaps. Except for N-terminal determinations of some of the tryptic and thermolytic peptides, all the sequences were determined by the DABITC/PITC double-coupling method [22], which offers the advantage of very high sensitivity and direct discrimination between acidic amino acids and their amides. Leu and Ile derivatives were identified chromatographically [23] or by back hydrolysis [17]. The DABITC/PITC method allowed routine identification of 5–15, and occasionally >20 residues, with only 4–20 nmol peptide. Assignment of the amide groups was done directly on the basis of the DABITC-degradation. Gln/Glu discrimination was confirmed by the specificity of the *Staphylococcus* protease. The aspartic acid residues identified were in agreement with the results obtained from the partial cleavage in dilute acid.

In the early phases of the work the cysteine-containing tryptic peptide T24 was not detected because the yield was low as is generally the case with cysteine peptides. Moreover, the peptide's N-terminal cysteine resulted in very low fluorescence after spraying with fluorescamine. This peptide was finally isolated as the *S*-[$^{14}$C]carboxamidomethylated derivative, but could not be degraded by the Edman procedure because of a probable cyclization reaction of the N-terminal carboxamidomethylated residue. This reaction has been reported [24] and was also observed by us during the sequence determination of rat prostatic binding protein [25].

Cyclization of the N-terminal glutamine in the tryptic peptides T5 and T7 gave similar problems if

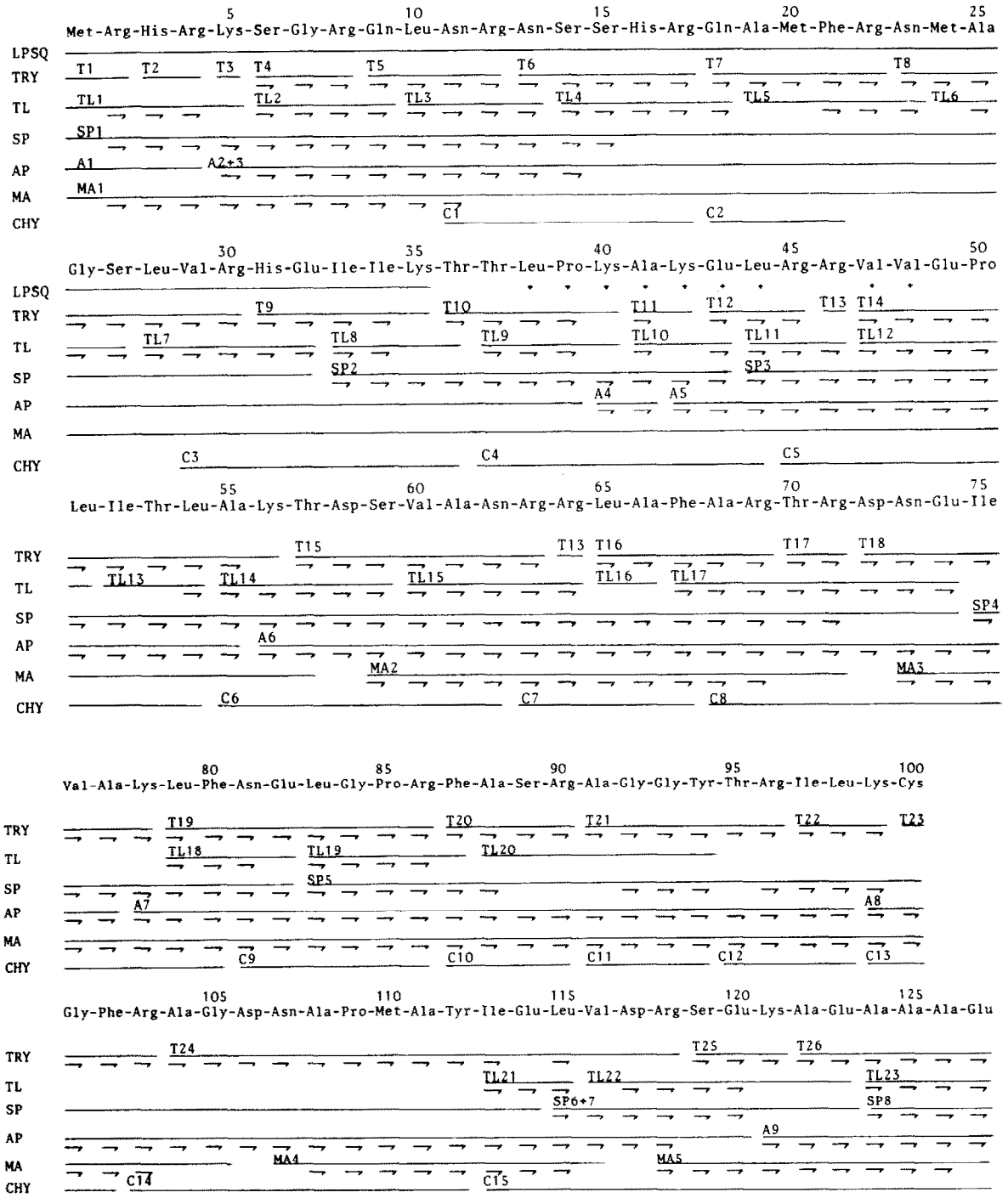



Fig. 1. Primary structure of protein L17. Sequence data on individual peptides are indicated as follows: →, sequenced with the double-coupling procedure [22]; LPSQ, sequenced on an improved Beckman sequencer. TRY, TL, SP, AP, MA and CHY indicate peptides derived from cleavage with trypsin, thermolysin, *S. aureus* protease, *A. mellea* protease, mild acid and chymotrypsin, respectively.

these peptides were eluted from the thin-layer sheets with 50% acetic acid. However, both peptides could be sequenced after elution with 20% pyridine in water. The sequence of the amino acids around the single cysteine residue was determined by the DABITC-Edman degradation of radioactive peptides from other digestions, i.e., peptide 8 from *Armillaria* protease and a peptide resulting from further cleavage with chymotrypsin of a large radioactive peptide obtained by dilute acid cleavage. In both these peptides, lysine was N-terminal, and *S*-carboxamidomethylcysteine was located in the second position. No cyclization was observed in the corresponding degradations. The residue was positively identified, both as its coloured thiohydantoin-derivative and by the distribution of radioactivity in the serial extracts.

All tryptic peptides were isolated and their positions in the sequence determined by overlaps mainly from peptides obtained by *Staphylococcus* and *Armillaria* proteases and by dilute acid hydrolysis (fig. 1). The alignment of the peptides in the N-terminal region was independently performed by automatic liquid-phase degradation of the protein up to position 35 [26]. Beyond this position, the alignment of the peptides in fig. 1 is in agreement up to position 50 with the results obtained from these sequencer runs. The identification of the released PTH-amino acid derivatives was made by thin-layer chromatography, back hydrolysis and by mass spectrometry. The residues at positions 38–44 and 47–49 were positively identified. Overlaps not aligned by the *Staphylococcus* and *Armillaria* protease peptides were unambiguously identified from thermolytic and chymotryptic peptides.

The cyanogen bromide fragments could be separated into 3 peaks by gel filtration on Sephadex G-25. Amino acid analysis showed the first peak, eluted at approximately the exclusion volume, to contain both intact L17 and the N-terminal part up to the methionine residue at position 110. Peaks 2 and 3 contained peptides corresponding to positions 2–20 and 111–127, respectively. The amino acid composition of these fragments was in agreement with the proposed sequence. They were not sequenced because the complete sequence had already been derived from the other fragments. However, the larger fragments from peak 1 were further digested with *Armillaria* protease and by dilute acid cleavage to check the proposed sequence alignment. The results of the carboxypeptidase digestions confirmed the accumulation of several alanine and glutamic acid residues at the C-terminal end of the protein chain and the lysine at pos. 121.

Full details of the sequence determination will be described elsewhere.

### 3.2. *Characteristics of the sequence*

As derived from the sequence presented in fig. 1 protein L17 has a $M_r$ 14365 and its amino acid composition is: $Asp_4$, $Asn_7$, $Thr_6$, $Ser_7$, $Glu_9$, $Gln_2$, $Pro_4$, $Gly_7$, $Ala_{17}$, $Val_6$, $Met_4$, $Ile_6$, $Leu_{11}$, $Tyr_2$, $Phe_5$, $His_3$, $Lys_8$, $Arg_{18}$, $Cys_1$. Tryptophan is not present. The total number of residues is 127. These results compare very well with the amino acid composition determined after hydrolysis of the protein.

The sequence has some noteworthy features:

(i) An accumulation of alanine and glutamic acid residues at the C-terminus. The last 27 residues contain only 2 arginines and 1 lysine compared to 6 acidic residues. Therefore the C-terminus of this protein carries a negative charge at physiological pH;

(ii) A marked accumulation of basic residues in the N-terminal part of the molecule. Indeed, the first 46 residues contain all 3 histidines; furthermore, 4 lysines and 9 arginines, in both cases, 50% of the total number. In addition, this 46 amino acid stretch contains only 2 glutamic acid residues. Its resulting net charge at physiological pH must therefore be strongly positive. This is especially true of the 5 first amino acids, which, except for the N-terminal methionine, are all basic. The sole cysteine residue is at position 100, and it is situated in a part of the molecule which predominantly contains apolar and aromatic residues, including one of the two tyrosines (pos. 94) and a phenylalanine residue (pos. 102);

(iii) L17 contains 4 methionine residues. One of these is the N-terminal residue, 2 are close together at positions 20 and 24, and the last one is located at position 110, next to the proline at pos. 109 and close to the second tyrosine residue at position 112.

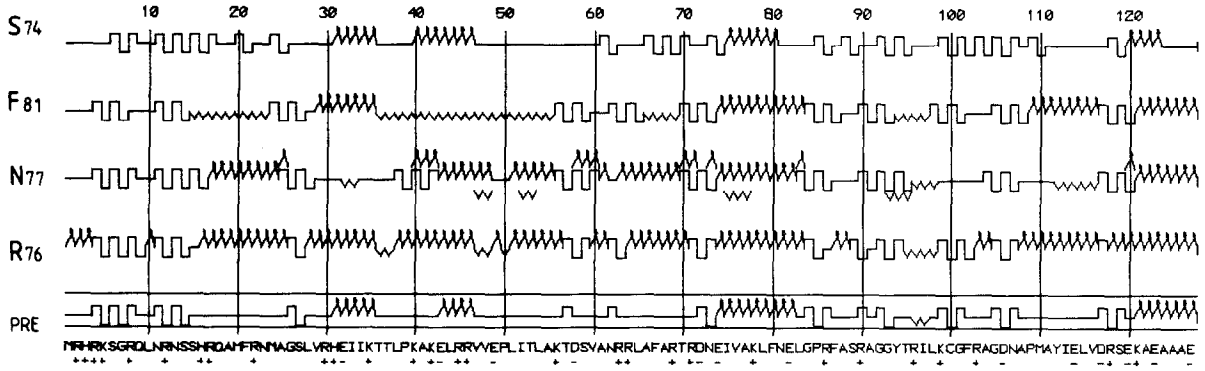



Fig. 2. Secondary structure predictions of protein L17. The calculations were made according to: [30], S; [28,29], F; [31], N; [32], R; ( ⩘ ) α-helix; ( ⩙ ) β-sheet; ( ⊓ ) turn structure; (——) random coil. In the line PRE 3 of 4 calculations gave predictions in agreement.

Table 1

Percentages of predicted secondary structure in protein L17

| Predictive method | α-Helix | β-Sheet | Turn | Unpredicted |
|---|---|---|---|---|
| Scheraga | 17 | 0 | 35 | 48 |
| Chou and Fasman | 12–28 | 26–43 | 35 | 11 |
| Nagano | 33–43 | 8–16 | 37 | 21 |
| Robson and Suzuki | 63 | 7 | 28 | 2 |

### 3.3. *Secondary structure predictions*

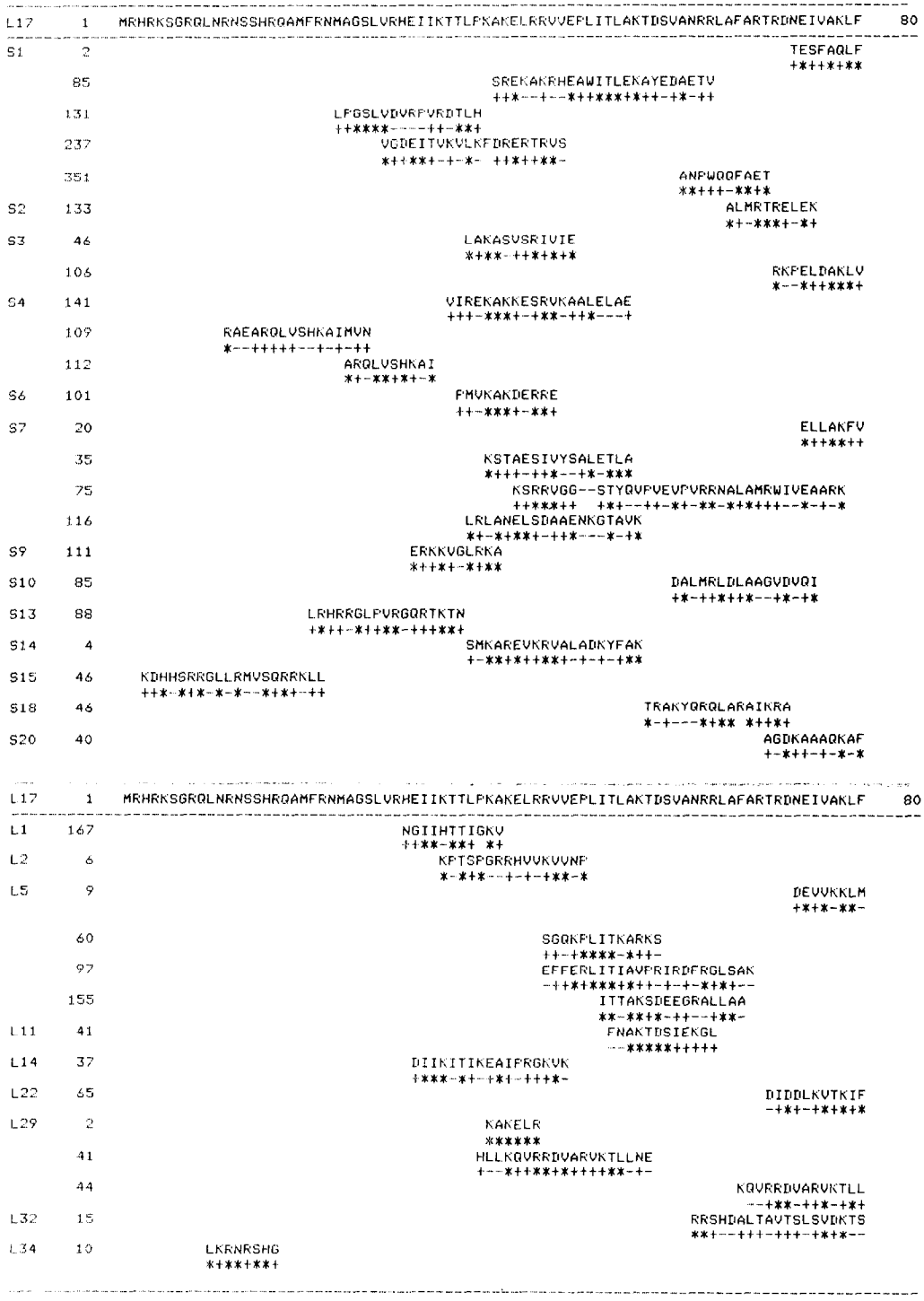
Possible secondary structural elements in protein L17 were calculated (fig. 2) employing 4 different methods [27]. In case of the prediction according to Chou and Fasman [28] a fully computerized version for the calculation has been used [29]. In table 1, the percentage values of calculated secondary structure given for the different predictions are presented. The rules of Chou and Fasman gave ambiguous results for positions 20–23, 29–35,

```
                1                                  10                                  20
E. coli   Met-Arg-His-Arg-Lys-Ser-Gly-Arg-Gln-Leu-Asn-Arg-Asn-Ser-Ser-His-Arg-Gln-Ala-Met-
B. stear.                     Ser-Tyr-Arg-Lys-Leu-Gly-Arg-Thr-Thr-Ser-Gln-Arg-Lys-Ala-Leu-

               21                                  30                                  40
E. coli   Phe-Arg-Asn-Met-Ala-Gly-Ser-Leu-Val-Arg-His-Glu-Ile-Ile-Lys-Thr-Thr-Leu-Pro-Lys-
B. stear. Leu-Arg-Asp-Leu-Ala-Thr-Asp-Leu-Ile-Ile-Asn-Glu-Arg-Ile-Glu-Thr-Thr-Glu-Ala-Arg-

               41                                  50                                  60
E. coli   Ala-Lys-Glu-Leu-Arg-Arg-Val-Val-Glu-Pro-Leu-Ile-Thr-Leu-Ala-Lys-Thr-Asp-Ser-Val-
B. stear. Ala-Lys-Glu-Leu-Arg-Ala-Val-Ile-Glu-Lys-Met-Ile-Thr-Leu-Gly-Lys-Arg-Gly-Asp-Leu-

               61                                  70                                  80
E. coli   Ala-Asn-Arg-Arg-Leu-Ala-Phe-Ala-Arg-Thr-Arg-Asp-Asn-Glu-Ile-Val-Ala-Lys-Leu-Phe-
B. stear. His-Ala-Arg-Arg-Gln-Ala-Ala-Ala-Phe-Ile-Arg----------------------Lys-Leu-Phe-

               81                                  90                                 100
E. coli   Asn-Glu-Leu-Gly-Pro-Arg-Phe-Ala-Ser-Arg-Ala-Gly-Gly-Tyr-Thr-Arg-Ile-Leu-Lys-Cys-
B. stear. Ser-Asp-Ile-Ala-Pro-Arg-Tyr-Gln-Asp-Arg-Gln-Gly-Gly-Tyr-Thr-Arg-Ile-Met-Lys-Leu-

              101                                 110                                 120
E. coli   Gly-Phe-Arg-Ala-Gly-Asp-Asn-Ala-Pro-Met-Ala-Tyr-Ile-Glu-Leu-Val-Asp-Arg-Ser-Glu-
B. stear. Gly-Pro-Arg-Arg-Gly-Asp-Gly-Ala-Pro-Met-Val-Ile-Ile-Glu-Leu-Val CT.

              121
E. coli   Lys-Ala-Glu-Ala-Ala-Ala-Glu CT.
```

Fig. 3. Sequence comparison of the homologous ribosomal proteins L17 of *E. coli* (this paper) and *Bacillus stearothermophilus* (M. Kimura, personal communication).

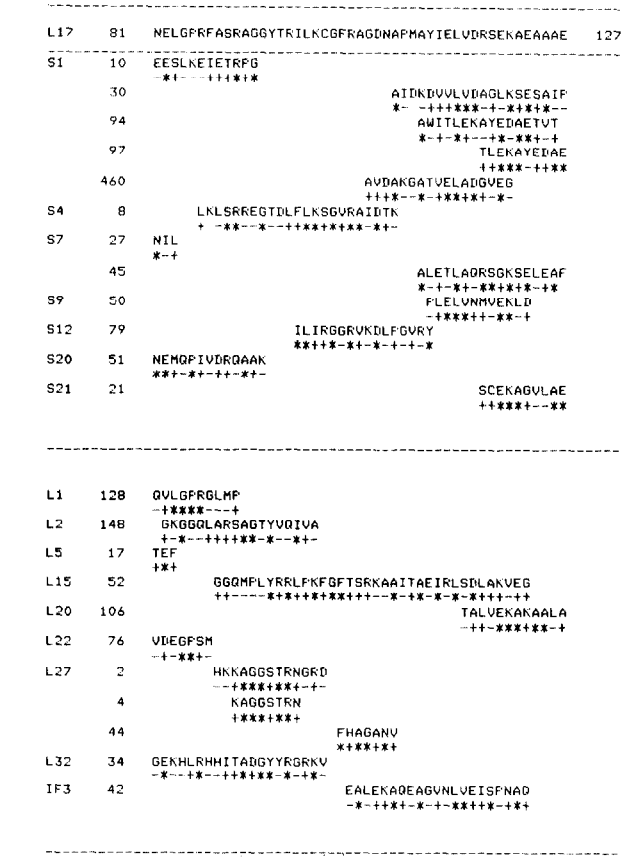
Table 2

Homologous sequence regions in protein L17 and other ribosomal proteins of *Escherichia coli*




(continued)

Table 2 (continued)



Searching of homologous sequence stretches was made by the aid of a computer programme for different length (8, 10, 12, 14, 16, 20, 24 and 30 residues). The accuracy was set to 35–50% identities among the two proteins compared; and of the remaining non-identical amino acids at least 35% had related codons differing in one or two nucleotides: (*) identical residues; (+) amino acids whose codons differ by 1 nucleotide; (–) amino acids whose codons differ by 2 nucleotides.

75–80 and 110–113 where both $\alpha$-helix and $\beta$-sheet were predicted with similar probability (not shown in fig. 2). According to these calculations, uniformly predicted helices were obtained for positions 74–82/83 and at the C-terminus from position 120/121. The protein has predicted extended structure for positions 15–23, 36–55 and 94–98, whereas turns are strongly predicted for the N-terminal region at positions 4–27, and at 56–59, 70–73, 84–87, 90–93, 99–110 and before the C-terminal helix.

### 3.4. *Comparison with sequences of other ribosomal proteins*

Computer programmes have been employed to compare the sequence of protein L17 with all other ribosomal structures. Among the different proteins from the *E. coli* ribosome several stretches of homologous structure were found. In table 2 some of these sequences are listed; the rest will be documented elsewhere. Noteworthy are regions of protein L17 which are similar to regions in protein L5, L15 and some of the other ribosomal proteins binding to the RNA (see table 2).

Strong homology has been found to a protein from *Bacillus stearothermophilus* ribosomes (M. Kimura, personal communication), and this protein is the homologue to protein L17 from *E. coli* (see fig. 3).

## ACKNOWLEDGEMENTS

We thank Dr H.G. Wittmann for providing samples of protein L17 and Dr V. Barkholt-Pedersen for a gift of *Armillaria mellea* protease. Some of the peptides have been isolated and sequenced by Ms H. van Bellegem. The technical assistance of Ms A.M. Ickroth, Mr A. Bosserhoff and the help with the design of search-programmes by Dr M. Dzionara is greatly acknowledged. We thank Mr K. Ashman for kindly reading the English version of the manuscript. This work was supported in part by the Belgian Nationaal Fonds voor Wetenschappelijk Onderzoek.

## REFERENCES

[1] Kaltschmidt, E. and Wittmann, H.G. (1970) Proc. Natl. Acad. Sci. USA 67, 1276–1282.
[2] Wittmann, H.G. (1974) in: Ribosomes (Nomura, M. et al. eds) pp. 93–114, Cold Spring Harbor Lab. Press, Long Island NY.
[3] Stöffler, G., Bald, R., Kastner, B., Lührmann, R., Stöffler-Meilicke, M., Tischendorf, G. and Tesche, B. (1979) in: Ribosomes: Structure. Function and Genetics (Chambliss, G. et al. eds) pp. 171–205, Univ. Park Press, Baltimore MD.
[4] Lake, J.A. (1979) in: Ribosomes: Structure, Function and Genetics (Chambliss, G. et al. eds) pp. 207–236, Univ. Park Press, Baltimore MD.
[5] Traut, R.R., Lambert, J.M., Boileau, G. and Kenny, J.W. (1979) in: Ribosomes: Structure, Function and Genetics (Chambliss, G. et al. eds) pp. 89–110, Univ. Park Press, Baltimore MD.

[6] Wystup, G., Teraoka, H., Schulze, H., Hampl, H. and Nierhaus, K.H. (1979) Eur. J. Biochem. 100, 101–113.
[7] Marquardt, O., Roth, H.E., Wystup, G. and Nierhaus, K.H. (1979) Nucleic Acids Res. 6, 3641–3650.
[8] Dohme, F. and Nierhaus, K.H. (1976) J. Mol. Biol. 107, 585–599.
[9] Sonenberg, N., Wilchek, M. and Zamir, A. (1973) Proc. Natl. Acad. Sci. USA 70, 1423–1426.
[10] Ghosh, N. and Moore, P.B. (1979) Eur. J. Biochem. 93, 147–156.
[11] Wittmann, H.G., Littlechild, J.A. and Wittmann-Liebold, B. (1979) in: Ribosomes: Structure, Function and Genetics (Chambliss, G. et al. eds), pp. 51–88, Univ. Park Press, Baltimore MD.
[12] Hindennach, I., Stöffler, G. and Wittmann, H.G. (1971) Eur. J. Biochem. 23, 12–16.
[13] Kaltschmidt, E. and Wittmann, H.G. (1970) Anal. Biochem. 36, 401–412.
[14] Vandekerckhove, J., Rombauts, W., Peeters, B. and Wittmann-Liebold, B. (1975) Hoppe-Seyler's Z. Physiol. Chem. 356, 1955–1976.
[15] Kamp, R., Yao, Z.-J. and Wittmann-Liebold, B. (1982) submitted.
[16] Lewis, W.G., Basford, Z.M. and Walton, P.L. (1978) Biochim. Biophys. Acta 522, 551–560.
[17] Peeters, B., Rombauts, W., Mous, J. and Heyns, W. (1981) Eur. J. Biochem. 115, 115–121.
[18] Wittmann-Liebold, B., Brauer, D. and Dognin, J. (1977) in: Solid Phase Methods in Protein Sequence Analysis (Previero, A. and Coletti-Previero, M.A. eds) pp. 219–232, Elsevier Biomedical, Amsterdam, New York.
[19] Wittmann-Liebold, B. (1980) in: Polypeptide Hormones (Beers, R.F. jr and Bassett, E.G. eds) pp. 87–120, Raven, New York.
[20] Wittmann-Liebold, B. (1981) in: Chemical Synthesis and Sequencing of Peptides and Proteins (Liu, T.Y. et al. eds) pp. 75–110, Elsevier Biomedical, Amsterdam, New York.
[21] Vandekerckhove, J., Rombauts, W. and Wittmann-Liebold, B. (1977) Hoppe-Seyler's Z. Physiol. Chem. 358, 989–1002.
[22] Chang, J.Y., Brauer, D. and Wittmann-Liebold, B. (1978) FEBS Lett. 93, 205–214.
[23] Yang, C.Y. (1979) Hoppe-Seyler's Z. Physiol. Chem. 360, 1673–1675.
[24] Edman, P. (1970) in: Protein Sequence Determination (Needleman, J.B. ed) p. 245, Springer-Verlag, Berlin.
[25] Peeters, B., Heyns, W., Mous, J. and Rombauts, W. (1982) Eur. J. Biochem. 123, 55–62.
[26] Wittmann-Liebold, B., Geissler, A.W. and Marzinzig, E. (1975) J. Supramol. Struct. 3, 426–447.
[27] Dzionara, M., Robinson, S.M.L. and Wittmann-Liebold, B. (1977) Hoppe-Seyler's Z. Physiol. Chem. 358, 1003–1019.
[28] Chou, P.Y. and Fasman, G.D. (1978) Adv. Enzymol. Relat. Areas Mol. Biol. 47, 45–148.
[29] Rawlings, N., Ashman, K. and Wittmann-Liebold, B. (1982) submitted.
[30] Burgess, A.W., Ponnuswamy, P.K. and Scheraga, H.A. (1974) Israel J. Chem. 12, 239–286.
[31] Nagano, K. (1977) J. Mol. Biol. 109, 251–274.
[32] Robson, B. and Suzuki, E. (1976) J. Mol. Biol. 107, 327–356.